MULTIMEDIA UNIVERSITY®

STUDENT ID NO

MULTIMEDIA UNIVERSITY

FINAL EXAMINATION

TRIMESTER 2, 2016/2017

HPB2031 – BIOINFORMATICS ALGORITHM I

(All sections / Groups)

09-MARCH-2017
9:00AM-11:00AM
[2 hours]

INSTRUCTIONS TO STUDENTS

1. This question paper consists of 3 pages, including this cover page.
2. You are required to attempt all questions. All questions carry equal 10 marks.
3. Please write all your answers in the Answer Booklet provided.
4. You may use a calculator.

Question 1

a) What are biological databases? Briefly explain the types of biological databases available and give the example database from each type. [2 marks]

b) What are the difference between the Smith-Waterman and Needleman-Wunsch algorithm used in sequence alignment? Give an example of when each would be used. [4 marks]

c) Briefly describe the following
   i. Gap Penalty
   ii. Local Alignment
   iii. Global Alignment
   iv. Optimal Alignment [4 marks]

Question 2

a) Calculate a BLOSUM matrix for the following sample:

Given a block of 4 A's, 1 B and 1 C

...A...
...A...
...B...
...A...
...C...
...A...

List all the amino acids pairs (substitutions - F_{xy}) and calculate the probability of each (P_{xy}). Calculate the expected frequency of A, B, C (q_x). Calculate the odds scores as observed probabilities / expected probabilities. Convert the score to log-odd score using base 2 with a scaling constant of 2. Assume no evolutionary history is known. [4 marks]

b) How is de Brnijin graph being implemented in genome assembly? What is the advantage of de Brnijin graph algorithm in the genome assembly as compared to overlap-layout consensus? [3 marks]

c) How do you improve genome assembly using de Bruijin graph-based assembler? [1 mark]

d) Under what situations where a genome draft instead of complete genomes is generated? [2 marks]

Continued...

Question 3

a) What is the relationship and differences between protein-protein interaction network and metabolic networks? [2 marks]

b) Why is the super matrix phylogenomics approach is not fully fitted into the context of evolution? [2 marks]

c) What are the key concepts for distances in UPGMA and Neighbour Joining phylogenetic trees? [4 marks]

d) In the effort to annotate a newly sequenced genome, there are two homologs identified instead of one as observed in the reference genome from the same species. What would you do to suggest whether the paralog is the result of gene duplication or horizontally transferred from different genome? Explain. [2 marks]

Question 4

a) What are the differences between supertree, supermatrix and bootstrap approaches in improving the reliability of phylogeny inference? [3 marks]

b) Reconstructs phylogenetic tree using UPGMA approach based on the dataset below. Show the steps and draw the resulting tree with the branch length. [3 marks]

	A	B	C	D	E
A	0	9	8	5	3
B		0	3	7	6
C			0	10	10
D				0	5
E					0

c) In the context of Pathways, what is a PGDB? What are the components (tiers) that exist in PGDBs? [4 marks]

END OF PAPER.